Nucleic Acids Research, Vol. 18, No. 6

A factor binding GATAAG confers tissue specificity on the promoter of the human ζ-globin gene

Paul Watt, Peter Lamb, Lyndall Squire and Nicholas Proudfoot
Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Received January 25, 1990; Accepted February 5, 1990

ABSTRACT

We describe the characterisation of cis-acting sequences which control the tissue specific expression of the human ζ globin gene. An extensive search for enhancer sequences in the vicinity of this gene proved negative. Instead our data demonstrate that the minimal promoter of the ζ gene is itself tissue specific. Sequences close to and possibly including the −100 CACCC and −70 CCAAT boxes display some erythroid specificity. However the principal tissue specific element is a GATAA sequence at −120 directly adjacent to the minimal promoter. Specific deletion of GATAA reduces ζ promoter activity 5 fold in erythroid but not non-erythroid cells. We also demonstrate that an erythroid specific factor binds to this GATAA sequence. Furthermore this factor forms a complex with the transcription factor CP1 which we show interacts with the ζ CCAAT box. We present evidence that the ζ GATAA binding factor is equivalent to GF1 recently purified and cloned by Tsai et al [1]. The erythroid specific GATAA sequence has been found in the promoters and enhancers of a number of erythroid specific genes. Similarly we show here that the ζ globin gene relies on a GATAA sequence in its promoter to specify its expression in erythroid cells.

INTRODUCTION

Human embryonic globin genes ζ and ε are expressed in the blood islands of the yolk sac from about week 3 of development until they are replaced by the expression of the foetal globins α and γ, at around the 6th and 7th weeks respectively. Their exclusive expression in erythroid tissue makes the globin genes an ideal model system for studying the molecular basis of tissue specificity [2]. The transcription of the globin genes is also stage specific, with the expression of embryonic globins being followed by foetal globins before the eventual production of the adult globins α and β.

The availability of cell lines [3−5] expressing globins from each of these stages has allowed the identification of a number of DNA signals associated with regulation of the β-like globin gene cluster. By the transfection of these genes into erythroid cell lines as well as in transgenic mice experiments, promoter and enhancer sequences have been defined for the γ and β globin genes [6]. More recently a distant regulatory element positioned about 50 Kb 5′ to ε globin has been identified called the dominant control region (DCR). This element functions as a potent erythroid specific enhancer, but unlike other enhancers, confers position independent and copy number dependent expression on linked genes [7]. These novel characteristics of the β cluster DCR suggest that it may function specifically to expose the higher-order chromatin structure associated within the β globin gene cluster in erythroid tissue. The more local promoter and enhancer elements associated with the different β-like globin genes may then act to control the stage specific expression of the ε, γ and β globin genes during erythroid development.

In contrast to the β globin cluster, relatively little progress has been made to identify key regulatory elements in the human α globin gene cluster. The adult α globin gene, although erythroid specific *in vivo*, displays a surprising level of activity when transfected into non erythroid cells, even in the absence of a closely linked viral enhancer [8,9]. In contrast, the ζ globin gene behaves in a highly tissue specific manner and only displays transcriptional activity when transfected into erythroid cell lines [10] (such as K562) but not in non erythroid cells such as Cos 7 or HeLa [11]. By analogy with the β globin gene cluster, it seems likely that the α cluster may similarly possess a DCR sequence possibly some distance from the ζ and α globin genes. However it also seems likely that the observed tissue specificity of ζ globin might be mediated by more local enhancer or promoter signals as has been shown for the γ and β globin genes.

We therefore have carried out a detailed investigation of sequences in and adjacent to the ζ globin gene to identify such regulatory elements. We first made an extensive search for an enhancer element near ζ globin. Finding none, we confined our attention to the ζ promoter. The general structure of this promoter is shown in figure 1. We have used *in vitro* mutagenesis to study the region of the promoter from −116 to the ζ-cap site. This region contains the GATAAG, CACCC, CCAAT and TATA consensus sequences commonly found in globin genes. We present *in vivo* and *in vitro* evidence that an erythroid factor binding to the the GATAAG motif of the promoter has a functional role in stimulating tissue-specific transcription of the ζ-globin gene. Similar GATAAG elements have been found in the promoters and enhancers of other globin and non-globin genes which are erythroid specific [12−13]. A tissue-specific but not stage-specific factor from the nuclei of erythroid cells, has been found to bind to this element *in vitro*. Such an erythroid factor has been described as Eryf1 in chickens [13] and GF1/NFE/NFE1 in humans and mice [14−16]. These factors are probably very similar [14]. We demonstrate that other tissue

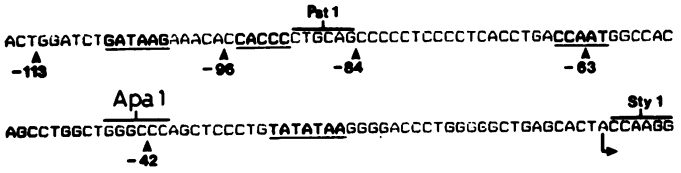


Figure 1
The sequence of the human ζ-globin gene promoter. The sense strand is shown from −116 to +7 (Cap site at +1, indicated by arrow). Unique restriction sites are shown above the line, together with the underlined conserved promoter elements GATAAG, CACCC, CCAAT and TATA. Below the line are shown the endpoints of various 5′ deletion constructs. The constructs used in transfection experiments, were made by fusing the CAT gene of vector pCATO, to the Sty1 restriction site on the 3′ side of the ζ promoter cap site.

specific elements exist in the ζ-promoter, but that these alone are unlikely to account for the high level of expression of ζ-globin in early erythroid cells. Since it appears that ζ-globin lacks an endogenous enhancer, we suggest that the GATAAG motif in the ζ-globin promoter plays a dominant role in the expression of this gene in erythroid cells.

MATERIALS AND METHODS

Plasmid Constructions

All DNA manipulations were performed essentially as described in Maniatis et al. [17]. The CAT expression plasmid pCATO was made by ligating the BglII-BamHI fragment containing the CAT gene from the plasmid SVO BglII, which has a BglII site instead of a HindIII site 5′ of the CAT gene [18], into the BamH1 site of pUC119 such that the 5′ end of the CAT gene was adjacent to the Sma 1 site of the pUC119 polylinker. The plasmid pCATE was then formed by ligating the SV40 enhancer (nucleotides 101 to 346 of the SV40 sequence) into the blunt-ended BamH1 site 3′ of the CAT gene. pCATO and pCATE were the vectors into which fragments of ζ-globin promoter DNA were inserted to create pCATOZ and pCATEZ derivatives (see below).

The construction of a series of ζ/CAT fusion plasmids containing progressive truncations of the ζ-promoter, is described by Lamb et al [19]. The SV40 enhancer was subsequently removed from these constructs by subcloning the EcoRI fragments containing the ζ-promoter of these plasmids into pCATO which lacks an enhancer. The constructs used in the enhancer search were made by subcloning fragments from within and around the ζ-gene into the polylinker of pCATOZ vectors containing a short (from −113) or long (from −550) ζ-promoter. A generalized map of these constructs is shown in figure 2.

The base substitution mutants were made by synthesising complementary 39 base oligonucleotides containing consecutive 5 base transversion mutations. The annealed oligos have sticky SacI and PstI ends, allowing them to be cloned between the natural PstI site in the ζ-promoter and the SacI site in the pUC119 polylinker. The oligos have the half SmaI site and KpnI site adjacent to the sticky SacI end as is present in the pUC119 polylinker; hence the only differences between the wild type −116 ζ-CAT construct and the mutants are the 5 bp changes within the promoter (figure 3A). The sequence of the mutants was verified by sequencing; note that the mutant −111/−107 has a 1 bp deletion at position −109.

Cell culture, transfection and enzymatic assays

Putko cells [3] were grown in RPMI supplemented with 10% foetal calf serium, 2mM glutamine, 100μg/ml penicillin and 100U/ml streptomycin, and were split 1:5 every 2 days. HeLa cells and MEL cells [5] were grown in DMEM containing the same supplements and were split 1:3 (HeLa) or 1:5 (MEL) every 2 days.

Putko cells were transfected by electroporation. The day before transfection the cells were split 1:5. The following day the cells ($5\times10^6-10^7$) were pelleted and resuspended in 1 ml of Hepes Buffered Saline (pH7.1) containing 0.5% PEG 6000. DNA (50 μg of test plasmid, 5 μg of control plasmid) was added and the cells given a single shock of 1750 V at 25μF using a Biorad Gene Pulsar. MEL cells were transfected similarly but at 250V and 960μF capacitance. Cells were transferred immediately into half their original volume of media and harvested 40 hours later. HeLa cells were transfected using the standard calcium phosphate coprecipitation technique [20]. 20 μg of test plasmid and 5 μg of control plasmid were used per 90 mm plate. The precipitate was left on the cells for 12 hours, and the cells harvested 24 hours after its removal. Cellular lysates were prepared by sonication. The cleared lysates were assayed for β-galactosidase activity exactly as described by Herbomel et al [21] were then heated to 65°C for 5 minutes, and used in a standard CAT assay [18]. CAT assays were quantitated by cutting the chloramphenicol and 3-acetyl-chloramphenicol spots form the plate and counting them in a liquid scintillation counter. The% conversion figures so obtained were then divided by the% conversion of the background control plasmid pCATE or pCATO to give the relative CAT activities. The same preparation of background control plasmid was used throughout a particular set of experiments, as we have found that the level of background can vary between different preparations (and when different batches of cells are used). For this reason the relative CAT activities of two different sets of experiments cannot be directly compared.

RNAse protection assays

For analysis of CAT mRNA 5′ ends, cytoplasmic RNA was isolated from HeLa cells transfected with 50 μg of ζ-CAT plasmid and 5 μg of a plasmid (pβE) containing the human β-globin gene and SV40 enhancer [P.Lamb, unpublished].

RNA hybridisation probes were prepared by *in vitro* transcription of linearised plasmids [22] containing either the PstI-PvuII fragment from the -550 ζ-CAT plasmid (−84 in the ζ-promoter to +171 in the CAT gene) or the 350bp HinfI fragment that spans the Cap site of the human β-globin gene. Hybridisations contained 5×10^5 cpm of each probe and were incubated at 55°C for 16 hours. Digestion with RNAse A and T1 was at 20°C for 90 minutes. Resistant hybrids were ethanol-precipitated and run on a 6% denaturing polyacrylamide gel.

Oligonucleotides

Single stranded oligonucleotides (a kind gift from Ernie deBoer and Frank Grosveld) containing the human α-globin CP1 binding site (CTCCGCGCCAGCCAATGAGCGCCGCCC), a human β-globin NFE1 binding site (CGAGGCCAAGAGATATATC-TTAGAGGGAGT) and a point mutant of this sequence (CGAGGCCAAGAGATATATATTAGAGGGAGT), were each annealed to the their complementary strands at 50 ng/μl in 50mM KCL before being diluted to 3.33 ng/μl in TE and used as competitors in gel retardation assays (see below).

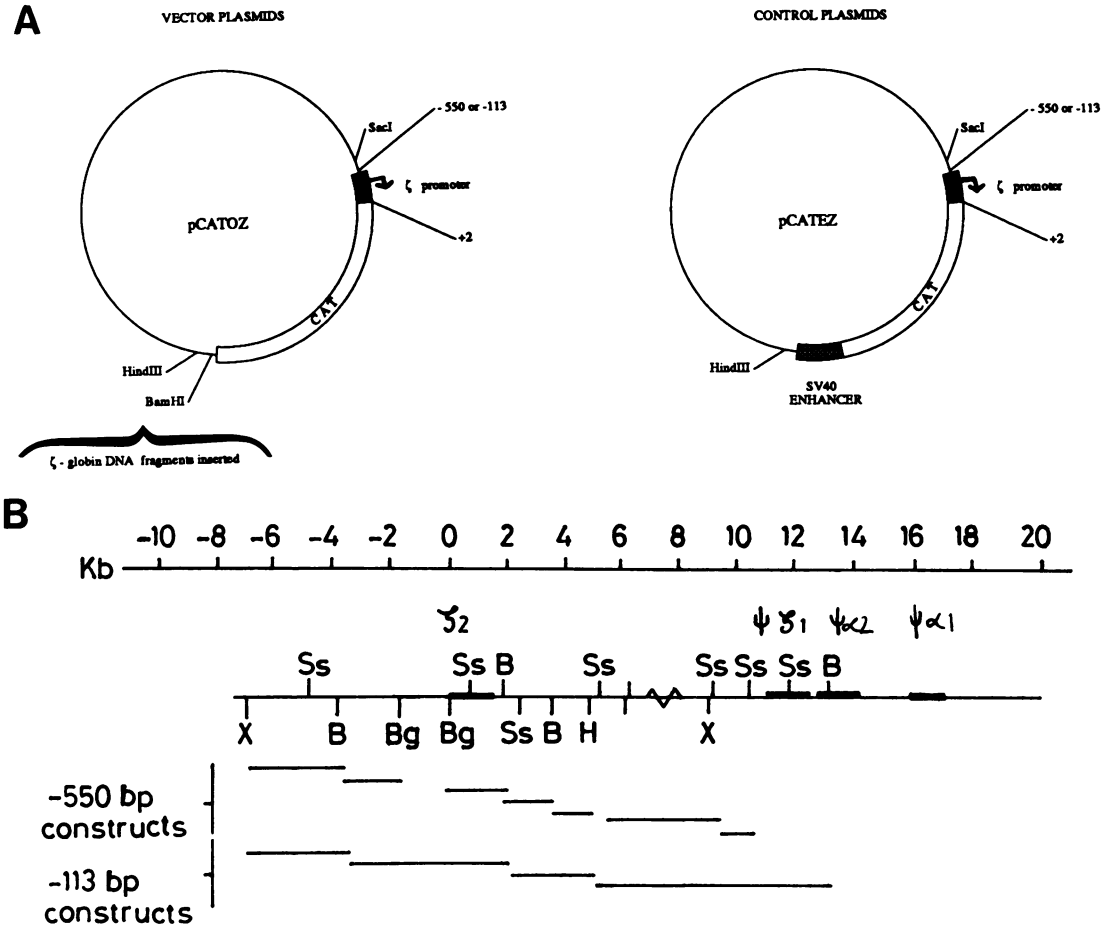


Figure 2

(A) Plasmid constructs used in enhancer search. The leftmost map shows the structure of the parent plasmid pCATOZ, from which vectors pCATZ550 and pCATZ113 were derived. pCATZ550 and pCATZ113 have 550 bp and 113 bp of ζ-promoter sequence respectively, upstream of the CAT gene. The fragments of ζ-globin DNA to be tested for the presence of an enhancer were inserted into the polylinkers of these vectors as shown (see figure 2B). The map on the right shows the structure of the parent plasmid pCATEZ from which the positive control plasmids pCATEZ550 and pCATEZ113 were derived. The position of the SV40 enhancer is indicated by the cross-hatched box. Only the relevant restriction sites are shown.

(B) The region of the human α-globin cluster, searched for the presence of an enhancer. The restriction map of the human α-globin cluster, shows a number of restriction sites and a scale in kilobases along the top. The abreviated restriction sites X, B, Ss, H and Bg correspond to XbaI, BamHI, SstI, HindIII and BglII, respectively. Below the map are shown the restriction fragments which were cloned into the vectors pCATZ550 and pCATZ113 (see figure 2A).

Preparation of Nuclear Extract

Nuclear extract was prepared from MEL and HeLa cells according to Dignam et al [23].

Gel Retardation Assays

The buffer for binding reactions contained 10mM Tris-Cl pH7.5 50mM NaCl, 1mM EDTA, 5% glycerol and 2 μg PolydIdC (Pharmacia). Each 20 μl binding-reaction contained 2−5000 cpm of labelled probe DNA and 0.5−2 μl of nuclear extract in binding buffer. The optimal concentration of extract preparation was determined by titration experiments for each new probe preparation. Each reaction was incubated at room temperature for 20 minutes, before loading directly (without adding dye) onto a 0.25×TBE native polyacrylamide gel (run at 350V for 2−3 hours).

In vitro Binding Competition

Double stranded competitor DNA was added to the nuclear extract in binding buffer (see above) and incubated at room temperature for 5 minutes. Then the probe DNA was added and the incubation continued for a further 20 minutes. Subsequent steps were as described above in 'Gel Retardation Assays'.

DNAse1 Footprinting Assay: Gel Retardation Protein: DNA complexes

Using MEL cell extract, a 5-fold scale-up of the standard DNA binding reaction described above was performed with 5mM $MgCl_2$ in the binding buffer. After 20 minutes of incubating the binding reaction, 1.5 Kunitz units of DNAse1 (Promega, RQ1) was added and the reaction continued for a further 2 minutes at room temperature before adding EDTA to 10 mM and loading a 4% native gel retardation gel (see above). (A control marker lane was treated in a similar way except a non-erythroid nuclear extract was used in the binding reaction.) The erythroid-specific low mobility band was isolated from the gel together with the band of unbound probe. These preparations were then phenol extracted and lyophilised, before resuspending in formamide loading dye. The two samples were then counted by liquid

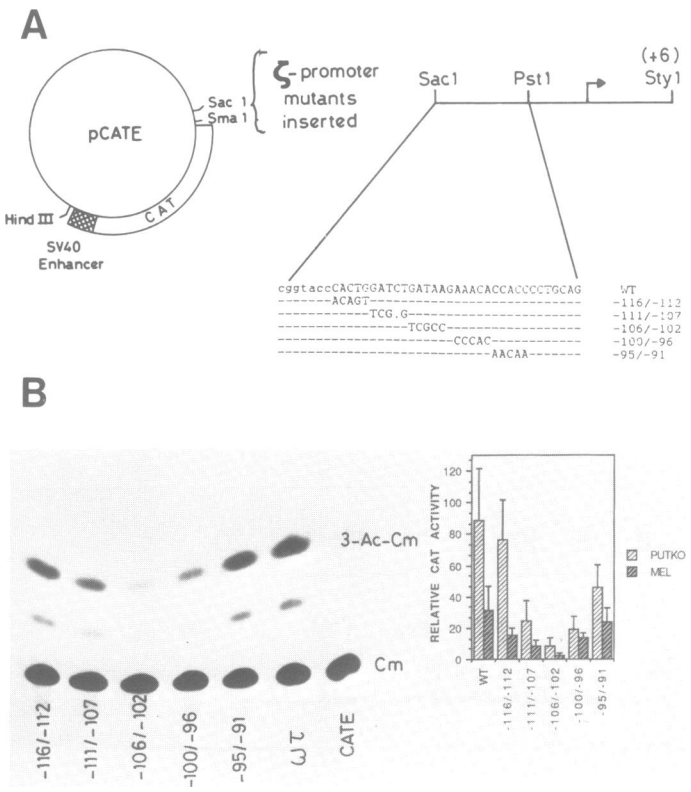


Figure 3

(A) A ζ-promoter base substitution mutants. The bases of the wild type (WT) sequence (sense strand) which have been altered are shown to the right of the vector plasmid in which they were inserted. Each of these mutations are in the context of a larger −116 to +6 promoter fragment.

(B) Sample CAT assay of the above base substitution mutants of the ζ-promoter in MEL cells. Crude whole cell Extracts were made from MEL cells which had been transfected with mutant plasmids or a promoterless control plasmid , together with a cotransfection control plasmid pIRV (see Methods). Extract concentration was equalized using an assay for expression of β-galactosidase from pIRV, and equivalent amounts of extract from each transfection were used in the CAT assay. The products of each assay were loaded onto the origin of a TLC plate. The lanes corresponding to the control plasmid pCATE and each deletion mutant are shown. CM denotes free ^{14}C chloramphenicol and 3-Ac-CM the acetylated derivative.

Relative CAT activities of the above base substitution mutants transfected into Putko cells and MEL cells are shown to the right of the CAT assay. The CAT activities shown are relative to the promoterless plasmid pCATO (=1 for Putko cells, =2 for MEL cells). The averaged results of four experiments are shown +/− s.e.m.

scintillation and equal counts loaded onto a 10% sequencing gel [24].

Solution Binding of DNA to Nuclear Extract

The same ζ promoter DNA probe (−116/−44, labelled on the antisense strand) was incubated with and without MEL nuclear extract in a binding reaction that had been preincubated (10 minutes) with polydIdC (as above). After 20 minutes incubation, DNase was added (as above) and incubated for 4, 8 and 10 minutes. The reactions were then purified as above and fractionated on an 8% sequencing gel [24].

RESULTS

Search For Tissue Specific Enhancer adjacent to the Human ζ-Globin Gene

The observed tissue specificity of human ζ-globin expression, suggested the presence of a tissue specific enhancer in or around the ζ-globin gene [10,11]. This possibility, together with our observation that the promoter of this gene is responsive to an exogenous enhancer [11,19], led us to search the DNA in the region of the ζ-globin gene for a natural endogenous enhancer.

We purified a range of DNA fragments spanning a region of 21kb in the vicinity of the ζ-globin gene, and subcloned them individually into restriction sites adjacent to a 550 base pair promoter fragment of ζ-globin which had been fused to the Chloramphenicol Acetyl Transferase (CAT) gene to create the vector plasmid pCATZ550 (see figure 2A). A fragment containing the SV40 enhancer was similarly introduced into pCATZ550 to produce the positive control plasmid pCATEZ550. The presence of a negative element in the ζ-promoter [19] could potentially make detection of an enhancer difficult, by dampening the transcriptional activity of the whole promoter. For this reason, a similar set of fragments from a region including the ζ-globin gene, was subcloned into sites adjacent to a truncated ζ-promoter which lacked the ζ negative element (this construct was called pCATZ113) (figure 2A). The location and size of the various fragments from the α-cluster which were subcloned, is shown in figure 2B. We then transfected these constructs into PUTKO (early erythroid phenotype) cells and assayed for any change in the activity of the ζ-promoter over vector background transcription, by measuring levels of CAT in cell extracts following a 48 hr period of expression. In these and in all other transfection experiments, a cotransfection control plasmid was also included to compensate for differences in transfection efficiency.

No enhancement over vector transcriptional levels was observed for any globin DNA fragment inserted near either the long or the short ζ-globin promoters. These data argue against the existence of a local ζ-globin enhancer. However, we did observe a considerable increase in CAT expression, when the SV40 enhancer was inserted into either vector pCATZ550 or pCATZ113 (data not shown).

It is formally possible that an endogenous enhancer of ζ-globin does exist outside the region which we have searched, or spanning a restriction site not covered by an overlapping clone. However we believe it improbable that an endogenous ζ-globin enhancer exists in this 21kb region. We therefore concentrated on defining cis-acting elements in the promoter of the ζ globin gene that may contribute to tissue specificity.

Effect of Base-Substitution Mutations on the activity of the ζ-Globin Promoter *in vivo*

To define promoter elements involved in expression from the ζ globin gene promoter, a series of base substitution mutants was made in the context of a ζ-globin promoter truncated to −116 and linked to the CAT gene. Five mutants were constructed which together provide a mutational scan of the region between −116 and −91 of the promoter. In each mutant, 5 consecutive base pairs are altered (figure 3A). The mutants, and a wild-type control construct were transfected into Putko cells and MEL cells (a mouse erythroid cell line) and the level of CAT expression from the ζ-globin promoter measured by CAT-assay (fig 3B). The experiment was repeated several times and the results quantitated (figure 3B). A significant reduction from the wild-type promoter activity was observed in mutants −101/−107, −106/−102, and −100/−96 in either cell type. In MEL cells, the mutant −106/−102 exhibited the largest (8-fold) effect while the neighbouring two mutations showed more modest 4-fold (−111/−107) and 2-fold (−100/−96) reductions in promoter

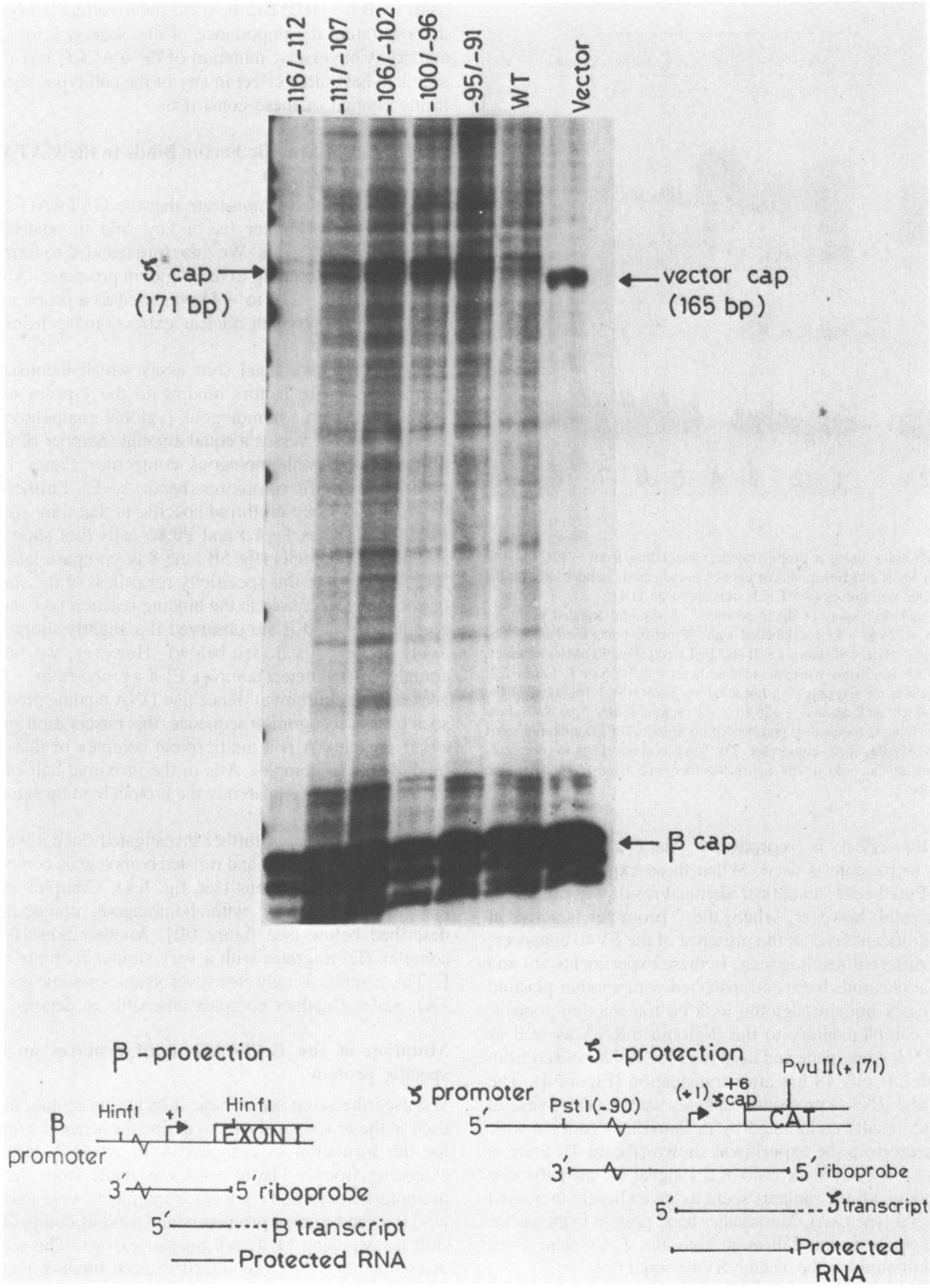


Figure 4

RNAse protection of RNA from HeLa cells transfected with the ζ-promoter base substitution mutants. Each mutant construct was transfected into HeLa cells. A promoterless vector construct was also transfected as a negative control. A cotransfection control pβE (see methods) was also included in each transfection and used to equalize the amount of RNA used in each sample of the RNAse protection experiment (described beneath the autoradiograph panel). The upper panel shows the resultant RNA run on a 6% denaturing polyacrylamide gel. The arrow marked ζ-cap points to the protected band of 171bp due to authentically initiated ζ-transcripts. The slightly smaller band of 165 bp represents readthrough transcripts into the CAT gene which initiate in vector sequences. The lower arrow indicates transcripts of the β-globin gene, originating from the cotransfection control plasmid pβE. The lower panel shows the positions of the ζ and β RNA probes and signals with respect to their gene maps.

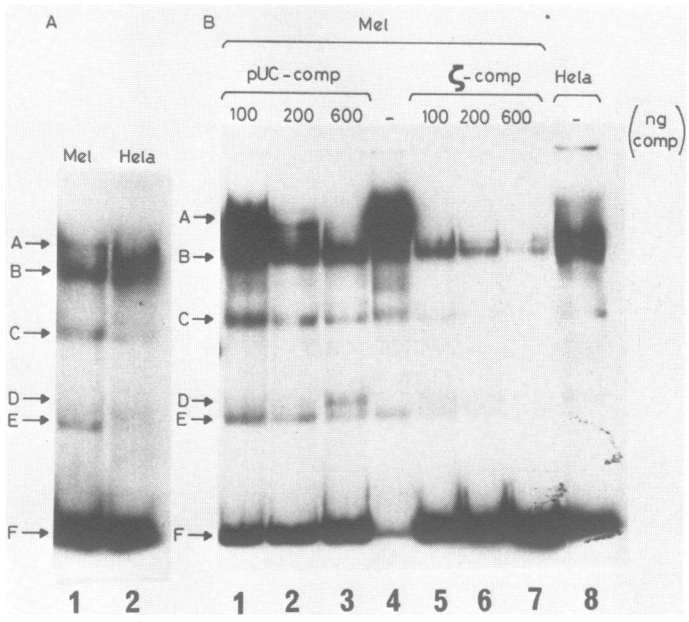


Figure 5
(A) A gel shift assay using a probe fragment extending from −126 to −42 incubated with MEL and HeLa nuclear extract as indicated. Letters indicate the different retarded complexes (A−E). F indicates free DNA.
(B) A similar gel shift assay of the ζ-promoter. A double stranded promoter fragment from −126 to −42 was labelled with ^{32}P before being used as a probe in gel shifts using MEL cell (lanes 1−7) and HeLa cell (lane 8) nuclear extract. Competitor DNA fragments were included in lanes 1,2,3,5,6 & 7. Lanes 1,2, and 3 show increasing amounts of a homologous EcoR1-Sty1 fragment of the ζ-promoter, which includes the −126 to −42 region. Lanes 5,6, & 7 show corresponding amounts (equivalent numbers of nucleotides) of a non-homologous (pUC118 HpaII,HaeIII digest) competitor. The bands corresponding to complexes A-E are shown on the right of the figure together with band F which is free (unbound) probe.

activity. This effect is reproduced when an independent preparation of plasmid is used. When these experiments were repeated in Putko cells, an almost identical result was observed.

In HeLa cells, however, where the ζ promoter is active at a low but significant level in the presence of the SV40 enhancer, a strikingly different result is seen. In these experiments w/t and mutant ζ-Cat plasmids were cotransfected with another plasmid which expresses human β-globin as a co-transfection control. Riboprobes complementary to the β-globin mRNA as well as to ζ-Cat mRNA were made and used to probe RNA isolated from the above HeLa cells 48 hrs after transfection (Figure 4). The transfection and RNAse protection analyses were repeated several times and the results quantitated by densitometry (not shown). As can be seen from the experiment shown (figure 4), there is no significant reduction in the ratio of β:ζ signal for any particular mutant, although all the mutants seem to have slightly increased levels over wild-type (w/t). The smaller band present in the vector lane, corresponds to read-through into the CAT-gene from transcripts initiating in the vector SV40 sequence.

The simplest explanation of these results is that cells of erythroid origin such as MEL cells and Putko cells (but not HeLa cells) contain a specific, positively acting factor which is able to bind sequences between −111 and −96, causing a stimulation of transcription.

This region includes the sequences GATAAG, a motif found in many genes expressed at high levels in erythroid cells. Interestingly, the mutation which disrupts the GATAAG sequence itself (−106/−102) has the most severe effect in erythroid cells demonstrating the importance of this sequence for ζ promoter activity. Conversely, mutation of the CACCC box (−95/−91) seems to have little effect in any of the cell types tested, at least in the context of these constructs.

An Erythroid-Specific Factor Binds to the GATAAG Motif of ζ-Globin *In Vitro*

The above results demonstrate that the GATAAG sequence of the ζ globin promoter has a key role in controlling tissue specificity of this gene. We therefore decided to look for tissue-specific factors binding to the ζ-globin promoter. A ζ promoter fragment from −126 to −44 was used as a probe in native gel retardation assays with nuclear extracts made from HeLa and from MEL cells [23].

Figure 5 shows a gel shift assay which identifies potential erythroid-specific factors binding to the ζ-promoter. Binding competition with a homologous ζ-globin competitor (figure 5B lanes 5, 6 & 7), versus a equal amount (number of nucleotides) of unlabelled nonhomologous competitor (lanes 1, 2 & 3), revealed 5 specific complexes (bands A−E). Furthermore, both bands A and E are erythroid specific in that they are present in MEL cells (lanes 1−4) and Putko cells (not shown), but are absent in HeLa cells (fig 5B lane 8 & compare lanes of figure 5A). We observe this specificity regardless of the concentration of nuclear extract used in the binding reaction (not shown). Both complexes A and E are observed if a slightly shorter (−116 to −44) probe is used (see below). However, we fail to detect complex A (but detect complex E) if a probe from −116 to −89 is used (data not shown). Since few DNA binding proteins require such a large recognition sequence, this observation suggests that either complex A is a multiprotein complex or that the protein binding site of complex A is in the proximal half of the −116 to −44 fragment. Apparently the protein binding site of complex E is distal to −44.

Complex C was not further investigated since it was not tissue specific (see figure 5A) and resisted homologous competition with some probe preparations (see fig. 6A). Complex B however, competes specifically with homologous competitor and is described below (see figure 6B). Another non-tissue-specific complex (D) migrates with a very similar mobility to complex E. This complex is only clearly resolvable on some gels (ie. figure 5A), making neither complex amenable to detailed study.

Mutation of the GATAAG motif removes an erythroid-specific protein

We used the same set of base substitution mutations that were used in the *in vivo* analysis, to define the sequence requirements for the formation of complex A *in vitro*. A labelled probe extending from −116 to −44 was made from the wild type promoter. Unlabelled competitor fragments were made from the wild type and mutant promoters and used as competitors in gel-shift assays using MEL cell nuclear extract. The same amount of competitor DNA was added to each binding reaction.

Figure 6A shows the result of such a competition experiment. The tissue specific complexes A and E are clearly competed away by a wild-type competitor (compare lanes 6 & 7) and by mutants −116/−112, −100/−96, and −95/−91 (compare lanes 1, 4 & 5, with lane 7). In contrast, competitors made from the −111/−107 and −106/−102 mutant promoters fail to compete the tissue specific bands effectively (compare lanes 2 & 3 with lane 7). Significantly, it is these two mutations which have the

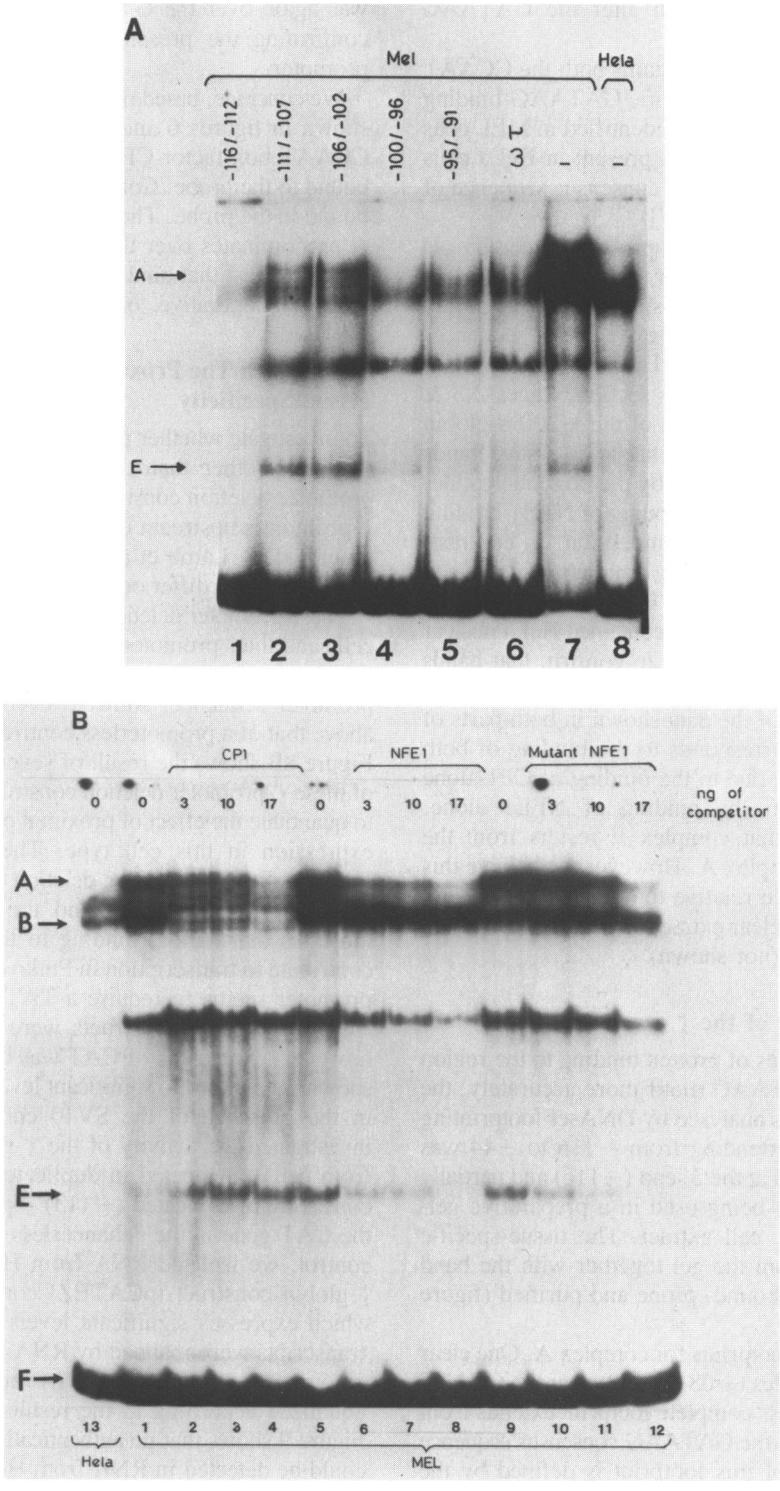


Figure 6

(A) Gel retardation assay using a wild type ζ-promoter probe and mutant competitors. A (−116 to −43) wild type probe was incubated with MEL cell nuclear extract (lanes 1 to 7) in the presence of unlabelled competitor fragments made from the base substitution mutants (lanes 1−5), or a homologous wildtype competitor (lane 6). The reaction of control Lane 7 contained no competitor DNA. The binding reaction of the rightmost control lane (8) contained the same probe incubated with HeLa cell extract without competitor, in order to show which complexes were tissue-specific. Equal amounts of competitor oligonucleotides were added to each binding reaction. Tissue-specific complexes A and E are indicated by arrows.

(B) A promoter fragment from −116 to −44 was used as a probe in binding competitions with MEL cell nuclear extract. The sequences of the three competitors are shown in the Materials and Methods section. Control lanes 1, 5 & 9 contain no competitor oligonucleotide. The competitor labelled CP1 (lanes 2,3 & 4), contains a CP1 binding site from the human β-globin promoter [27]. The competitor labelled NFE1 (lanes 6, 7 & 8) contains the high affinity NFE binding site from human β-globin promoter [16]. The mutant NFE competitor (lanes 10, 11 & 12) is identical except for a transition mutation within the NFE consensus (ATC to ATA). Equal amounts of each competitor were annealed and diluted to 3.33 ng/μl. Varying amounts of each competitor (shown above each lane) were preincubated with nuclear extract in the binding buffer prior to the gel retardation. The rightmost lane corresponds to a control containing no competitor and HeLa nuclear extract rather than MEL extract.

greatest effect *in vivo*, since they both alter the GATAAG consensus sequence.

The −116 to −44 probe fragment contains both the CCAAT and GATAAG motifs. An erythroid specific GATAAG-binding factor named NF-E1 has been previously identified in MEL cells [16,26]. Similarly a factor called CP1 is present in HeLa cells and MEL cells and binds to the CCAAT consensus sequence of the human α and β-globin genes [16,27].

In order to establish the identity of the proteins in complexes A and E, we used known binding sites for the proteins NFE and CP1 [16] in *in vitro* competition assays using a probe extending from −116 to −44 and MEL cell nuclear extract. The sequences of these oligonucleotides (a kind gift from De Boer and Grosveld) are shown in the Materials and Methods section. Lanes 2,3 & 4 of figure 6B, show increasing amounts of the −200 β-globin CP1 binding site as competitor in a gel retardation assay. Bands A and B are both significantly competed by the oligonucleotide. Lanes 6, 7 & 8 show that increasing amounts of NFE1 binding site removes complexes A and E by competition. In contrast, lanes 10, 11 & 12, which show increasing amounts of a mutated NFE1 site, fail to compete bands A and E as effectively as the wild-type NFE1 oligonucleotide. Lane 13 contained HeLa nuclear extract rather than MEL nuclear extract to confirm that bands A and E in lanes 1−12, are erythroid specific.

The simplest interpretation of the data shown in both parts of figure 6 is that complex A corresponds to the binding of both CP1 and NFE1, complex B is due to the binding of CP1 alone and complex E results from the binding of NFE1 alone. Alternatively, it is possible that complex E results from the proteolytic degradation of complex A. However we believe this is unlikely, since its abundance relative to complex A does not increase with the age of the nuclear extract or with the incubation time of the binding reaction (not shown).

DNAse I Footprint analysis of the ζ promoter

In order to characterise the sites of protein binding to the region of the promoter near the GATAAG motif more accurately, the DNA bound in complex A was analysed by DNAseI footprinting analysis. A DNA fragment extending from −116 to −44 was labelled on the antisense strand at the 3′ end (−116) and partially digested with DNAseI before being used in a preparative gel-retardation assay using MEL cell extract. The tissue-specific complex (A) was isolated from the gel together with the band (F) corresponding to free (unbound) probe and purified (figure 7A).

We observe two DNAse I footprints for complex A. One clear footprint lies between nucleotides (−68 & −59) near the CCATT box motif (−65). The other less complete footprint extends from −100 to −116 and aligns with the GATAAG consensus sequence at −107. The 5′ boundary of this footprint is defined by the beginning of the vector polylinker sequence at −117. We presume that the GATAAG footprint is partial due to the cross contamination of complexes A and B. Complex B does not contain NFE1.

To confirm the footprint analysis obtained for complex A we carried out additional footprint analysis of the same ζ promoter fragment in solution. The profiles of DNase digestion at increasing times were compared with and without MEL cell nuclear extract. As shown in figure 7B two footprints were again obtained. One contained the CCAAT sequence, but extended further 5′ suggesting the presence of an additional factor (factors) not detected in the gel retardation experiments. The other footprint was again over the GAATAG sequence between −100/−116 confirming the presence of an NFE1 binding site in the ζ promoter.

We conclude, based on the gel retardation and footprint analysis shown in figures 6 and 7, that complex A is composed of the CCAAT box factor CP1 and the GATAAG box factor (NF-E) bound to the probe. Complex E then corresponds to NF-E alone bound to the probe. The observation that the amount of complex A predominates over the amount of complex E (see figure 7), might suggest that binding to the CCAAT and GATAAG motifs may be cooperative, or mutually stabilizing.

Elements In The Proximal ζ-Globin Promoter contribute to Tissue Specificity

To investigate whether promoter elements contribute to cell-type specificity other than the GATAAG sequence, a series of 5′ promoter deletion constructs were made which placed truncated ζ-promoters upstream of the CAT gene. This deletion series is described by Lamb et al [19]. The deletion constructs used in this experiment differ only by the absence of the SV40 enhancer.

These promoter deletion constructs were transfected into Putko cells and their promoter activity determined by CAT assay. All of the truncated promoters with more than 63 bp of residual promoter sequence, showed levels of expression which were above that of a promoterless control plasmid (CATO) (Fig. 8A). Figure 8B shows the result of several independent transfections of these ζ-promoter deletion constructs into Putko cells, designed to quantitate the effect of proximal promoter elements in allowing expression in this cell type. The drop in expression levels following each successive deletion correspond to the removal of the CACCC-box (−95) and the CCAAT-box (−66). This indicates that factors binding to these sequences are likely to contribute to transcription in Putko cells. The minimal functional promoter seems to require a TATA-box and a CCAAT-box.

When the same constructs were transfected into HeLa cells however, expression of CAT was below background levels (not shown) in contrast to significant levels of expression in these cells in the presence of the SV40 enhancer (Figure 4). We also investigated the activity of the ζ promoter. RNA was isolated from cells transfected in duplicate with a construct (pCATZ), consisting of truncated (−113) ζ-promotor cloned upstream of the CAT gene in the enhancerless vector CATO. As a positive control, we isolated RNA from HeLa cells transfected with a ζ-globin construct (pCATEZ) containing the SV40 enhancer, which expresses significant levels of CAT. The 5′-ends of ζ-transcripts were mapped by RNAse protection analysis with the amount of cytoplasmic RNA added to each sample being equalized according to the results of a cotransfection control. Figure 9 shows that no authentically initiated ζ-globin transcripts could be detected in RNA from HeLa cells transfected with the enhancerless ζ-promoter plasmid (lane pCATZ) although there were such transcripts from the positive control plasmid (lane pCATEZ). We conclude from these experiments that the ζ-globin promoter does not direct detectable initiation in HeLa cells in the absence of an enhancer and that regulatory elements in the proximal region of the promoter contribute to cell type specificity.

DISCUSSION

In this paper we have set out to define the cis-acting sequences which control the tissue-specificity of ζ-globin gene transcription. It seems clear from an extensive search of the DNA near the

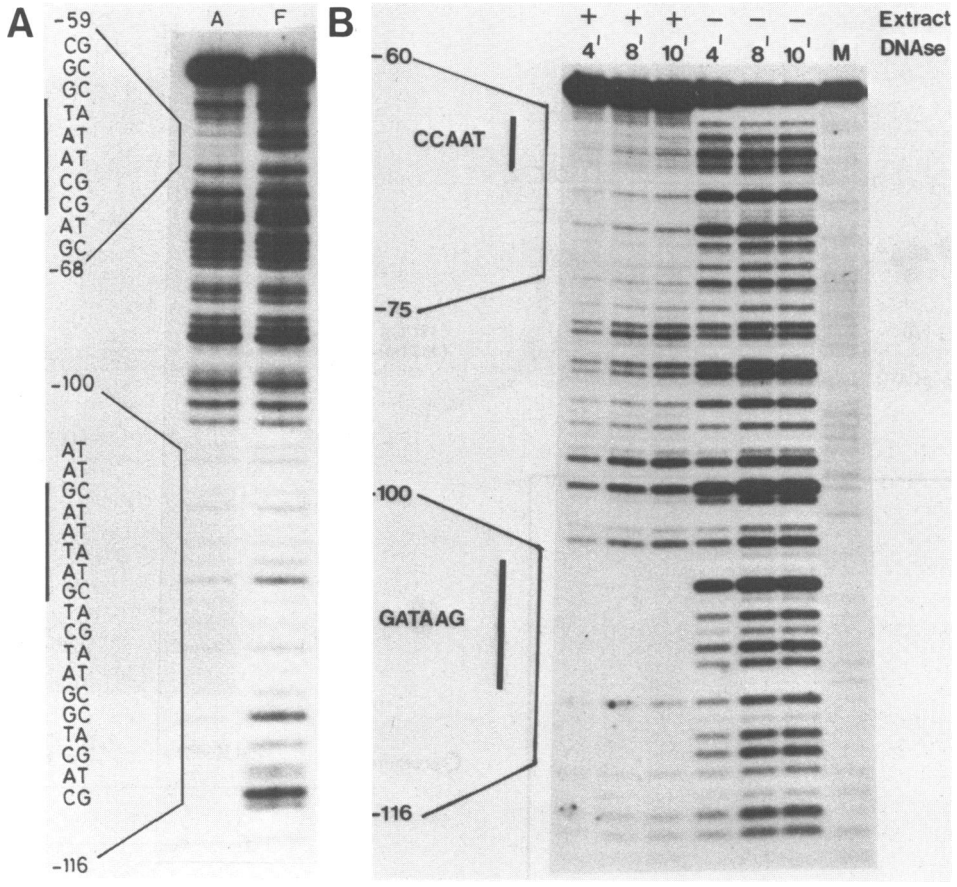


Figure 7

(A) DNAseI footprint analysis of tissue-specific complex A. The ζ-promotor region from −116 to −44 was incubated with MEL cell nuclear extract, labelled on the antisense strand and partially digested with DNAseI, beforebeing used in a gel shift assay. Complex (A) was isolated from the gel together with free probe (F). The DNA was purified and equal counts loaded onto a 10% sequencing gel. The coding sequence corresponding to each footprint is shown alongside. The distal boundary of the lower footprint is defined by the beginning of polylinker sequence.

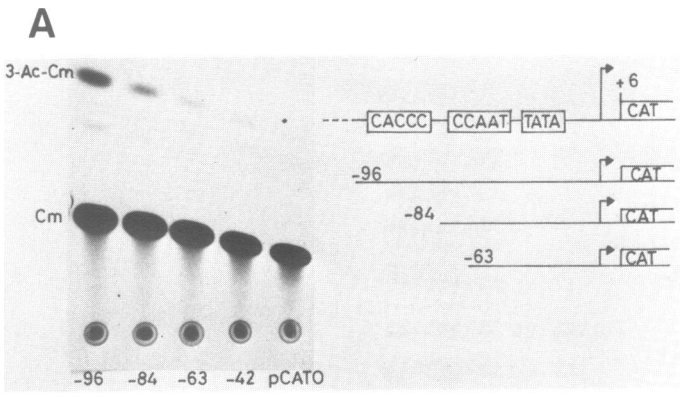
(B) DNAseI footprint analysis of the ζ promoter region −116 to −44 using solution binding of excess MEL cell nuclear extract. Increasing of DNAse incubation of the DNA probe are shown with and without nuclear extract. The two observed footprints are indicated together with the positions of the CCAAT and GATAAG sequences within these footprints.

ζ-globin gene (figure 2B) that there is no enhancer present. However when we studied the promoter in detail by mutagenesis, we found two regions contributing to erythroid cell-specific expression. The proximal region of the promoter which includes the CACCC and CCAAT boxes was sufficient to allow significant levels of authentically initiated transcription in Putko cells but not in HeLa cells. Deletions which successively remove each of these boxes cause a reduction in promoter activity in Putko cells. Since factors binding to these sequence motifs are not usually specific for erythroid cells, either other sequences in this region must be important, or novel factors binding to the CACCC or CCAAT motifs must be postulated. Mantovani et al [28] have reported such an erythroid cell-specific factor binding to the distal CCAAT box region of the human α-globin promoter. Such a factor could in principle be allowing the proximal ζ-globin promoter to function in Putko cells but not in HeLa cells. However, we find no evidence *in vitro* for such a tissue specific CCAAT box factor (see below). More work will be required to characterise further the precise location of these control elements. Since the effects of these proximal promoter deletions were small, it is clear that this proximal region alone cannot account for the observed tissue specificity of the ζ-globin promoter.

Evidence that the ζ-globin GATAAG box has a function *in vivo*

We made a series of base subtitution mutants in the distal half of a ζ promoter fragment extending to −116 bp and assayed the effects of these mutations in transient transfections of Putko, MEL and HeLa cells. Mutation of the sequence GATAAG itself or the surrounding nucleotides, has a drastic effect (up to 10 fold) on promoter activity in MEL and Putko cells (figure 3) but has no effect in HeLa cells which are only able to produce very low levels of ζ-transcripts despite the presence of the SV40 enhancer. This result implies that an erythroid factor binding the sequence GATAAG may be required for efficient transcriptional initiation of the ζ-globin promoter. There is another distal (−234) GATAAG box on the non-coding strand of the ζ-promoter. When a 51 bp sequence including this GATAAG sequence is deleted, a 2-fold reduction in promotor activity is observed [19] suggesting that it may also contribute to overall promoter activity.

Although several groups have reported such a DNA binding activity *in vitro* (for example, [29]), there are few reports providing direct functional evidence for the GATAAG element having an *in vivo* role in erythroid specific gene expression. However larger sequences containing this element in the



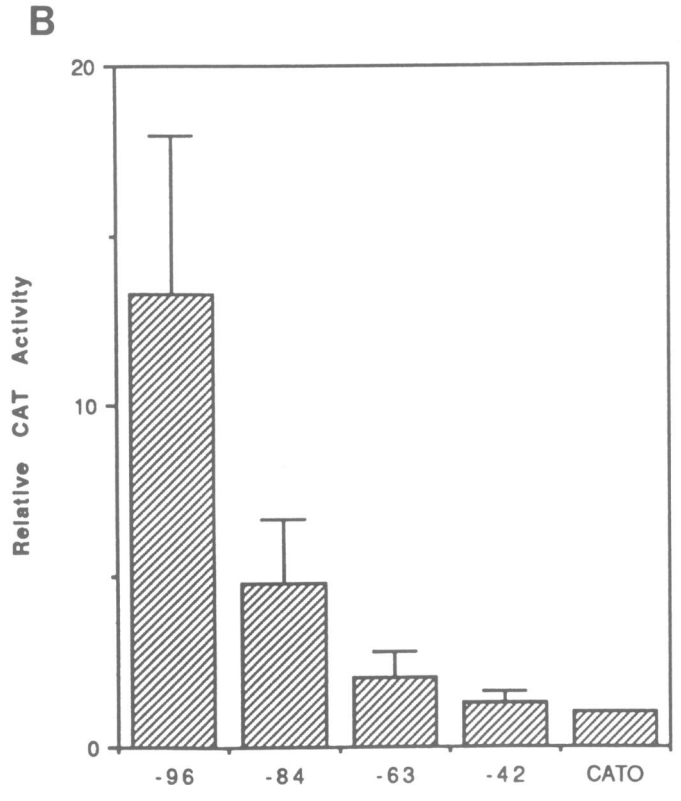


Figure 8
CAT assay of deletion mutants.(A) The lefthand figure shows a sample CAT assay of proximal 5′ deletion mutants of ζ-promoter in Putko cells. Crude whole cell Extracts were made from Putko cells which had been transfected with mutant plasmids or a promoterless control plasmid, together with a cotransfection control plasmid pIRV (see Methods). Extract concentration was equalized using an assay for expression of β-galactosidase from pIRV, and equivalent amounts of extract from each transfection were used in the CAT assay. The products of each assay were loaded onto the origin of a TLC plate. The lanes corresponding to the control plasmid pCATO and each deletion mutant are shown. CM denotes free ^{14}C chloramphenicol and 3-Ac-CM the acetylated derivative. The position of each of the deletion endpoints relative to conserved promoter elements is shown on the right.
(B) Relative CAT activities of proximal 5′ deletion mutants of ζ-promoter in Putko cells. The CAT activities shown are relative to the promoterless plasmid pCATO (=1). The averaged results of several experiments are shown +/− s.e.m.

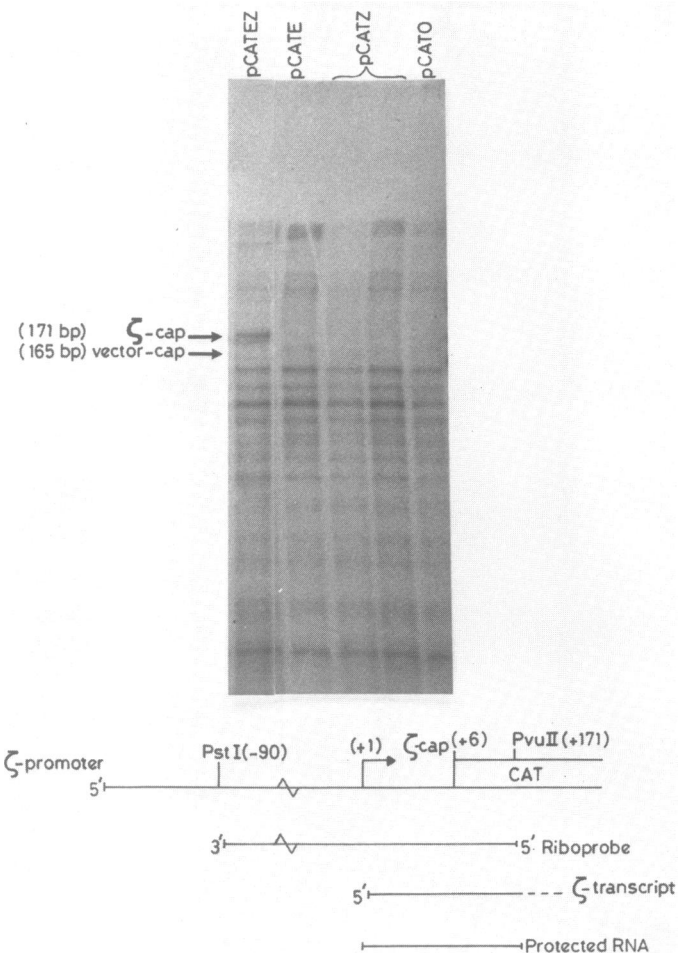


Figure 9
RNAse protection analysis of RNA from HeLa transfections of ζ-promoter deletion constructs. pCATZ was transfected into HeLa cells in duplicate. pCATEZ was also transfected as a positive control, together with corresponding negative controls pCATO and pCATE which lack the ζ-promoter. A cotransfection control pβE (see methods) was also included in each transfection and used to equalize the amount of RNA used in each sample of the RNAse protection experiment. The upper panel shows the resultant RNA run on a 6% denaturing polyacrylamide gel. The upper arrow points to the protected band of 171 bp due to authentically initiated ζ-transcripts. The lower band represents readthrough transcripts into the CAT gene which initiate in vector sequences. The lower panel shows the positions of the RNA probes and signals with respect to the gene map.

promoters of the α-globin and β-globin genes were shown to be responsible for inducible tissue-specific expression [12,16,30]. Moreover in the chicken β-globin enhancer, mutation of either half of an imperfect repeat containing the GATAAG consensus, resulted in a 5-fold reduction in enhancer activity [31].

Some of the best indirect evidence for GATAAG having a function *in vivo* comes from two natural mutations in the γ-globin promoter which are associated with hereditary persistance of foetal haemoglobin (HPFH) in which continued expression of γ-globin is observed as a result of a single base change [14]. In both of these HPFH mutations the affinity of factors for GATAAG consensus sequences are shown to be altered and may well be the cause of elevated γ-globin gene expression.

Evidence for factors binding to GATAAG and CCAAT motifs of ζ-globin

Since we have established an *in vivo* role for the GATAAG motif in the ζ-globin promoter, we decided to look for evidence that an erythroid-specific protein binds to this sequence *in vitro*. When a promoter fragment containing the GATAAG box, the CACCC

box, and the CCAAT box is used in gel retardation analysis, two erythroid specific complexes are observed, one of low mobility (complex A) and one of high mobility (complex E) (Figure 5). Both of these complexes are competed effectively by a wild type homologous promoter fragment, but fail to compete effectively with mutant promoter fragments where the GATAAG motif or surrounding sequences have been disrupted (figure 6A). These results indicate that each of these complexes contains an erythroid-specific protein. We then used competition with oligonucleotides corresponding to specific binding sites to define the proteins binding to this region more precisely. Based on these results, we suggest that the low mobility tissue specific complex contains both CP1 bound to the CCAAT box and NFE1 bound to the GATAAG motif. The high mobility tissue specific complex seems to correspond to the binding of NFE1 alone. To support this interpretation, we used DNAseI footprinting to show that the high mobility complex contained both CCAAT and GATAAG binding factors (figure 7). The two footprints which we observe for CP1 and NFE1, are similar to those described by others [14,16,29].

Sequence comparison of ζ-globin CCAAT box with CP1 sites of other genes

The CCAAT box in ζ-globin conforms well to the CP1 binding consensus of α-globin (Table I). Moreover, if we compare the homology of the ζ-globin CCAAT-box and eighteen flanking nucleotides with the archetype CP1 binding site of α-globin [27], we find that the ζ-globin CCAAT box has 10/23 nucleotides in common with the α-CCAAT box region, while the high affinity CP1 sites at −150 in β-globin and at −87 in α-globin share 9/23 and 7/23 nucleotides. Since we have no evidence for a tissue specific or heat stable CCAAT box factor, and the ζ-globin CCAAT box region bears no special relationship to CTF/NF1 [32] or CBP [33] binding sites, it is reasonable to assume that the factor which we observe binding to ζ-CCAAT is CP1. In support of this, the ζ-CCAAT-box complex has a similar mobility to CP1 [16] and is very sensitive to competition with a specific CP1 binding site (fig 7B).

Sequence comparison of ζ-globin GATAAG box with NF-E sites of other genes

A series of recognition sites for NFE1 (humans) and Eryf1 (chickens) have been used to define a consensus binding site sequence: A/C Py T/A ATC A/T Py [13,26]. The GATAAG box of human ζ-globin matches this sequence exactly. Comparison of the very strong Eryf1 binding site from the chicken α-globin promoter with the NFE sites of the human α-globin and human β-globin promoters, yields a 7/13 nucleotide match, while the GATAAG box of ζ-globin gives a 9/13 match (Table 1). Moreover, Eryf1 itself has been shown to bind to the promoter of the ζ-like gene of chickens (π-globin) [13].

Despite this strong indirect evidence for the identities of the two proteins which we observe in DNAse I footprinting, more conclusive proof will require gel-retardation assays using antibodies to purified CP1 and NFE1. In addition, the recent cloning of the murine homologue of NFE (GF1) [1] will allow the comparison of the tissue-specific complexes which we observe, with a GF1/ζ-GATAAG complex by gel retardation analysis.

A model for the activation of ζ-globin gene expression

It is tempting to speculate that the protein:DNA complex which contains factors simultaneously bound to both the GATAAG and CCAAT motifs has functional significance. Another group has observed an analogous complex on the promoter of the $^{A}\gamma$-globin promoter with a very similar relative mobility [15]. The GATAAG motif is often found close to either CACCC or CCAAT motifs suggesting an interaction between factors bound to these sites. Indeed the GATAAG motif placed in isolation upstream of a minimal (TATA box only) promoter, fails to activate erythroid-specific transcription, suggesting that the context in which the motif is placed is crucial to its function [12]. Likewise, the combination of NFE1 and high affinity CP1 binding sites, but not either of the sites alone, are capable of mediating the induction of transcription from a minimal (CACCC, proximal CCAAT & TATA box) β-globin promoter in MEL cells [16]. It is plausible that proteins bound to these two sites interact productively (possibly through DNA-looping) in the initiation of transcription. It would be interesting to see if there are any minimal spacing requirements for the GATAAG and other motifs in the promoter.

The wide distribution of the GATAAG motif on several genes throughout the α- and β-globin clusters suggests that it might play a coordinating role in the switching on of globin genes in the differentiation of the erythroid lineage. The discovery of the 'dominant control region' upstream of the β-globin cluster has allowed the position and copy-number independent expression of human globin genes in transgenic mice. Constructs containing this element may now be used in transgenic mice experiments to study the role of the various cis-acting regulatory DNA sequences in the developmental regulation of globin genes. Such studies will help to explain the mechanisms of interplay between tissue-specific proteins, and the individual elements of enhancers and promoters, in the formation of an active transcription complex.

ACKNOWLEDGEMENTS

P.W. is supported by a Commonwealth Scholarship from the Association of Commonwealth Universities, and P.L. by an MRC Training Fellowship. We gratefully acknowledge Ernie deBoer and Frank Grosveld for providing the CP1 and NFE1 oligonucleotides, and Emma Whitelaw for advice on the manuscript. This work was funded by an MRC project grant no.G8822153CB.

REFERENCES

1. S-F. Tsai, D.I.K. Martin, L.I. Zon, A.D. D'Andrea, G.G. Wong and S.H. Orkin (1989) Nature *339*, 446−451
2. F.S. Collins and S.M. Weissman (1984) Progress in nucleic acid research and molecular biology, vol. 31, pp. 315−421, Academic Press, New York.
3. G. Klein, J. Zeuthen, I. Eriksson, P. Terasaki, M. Bernoco, A. Rosen, G. Masucci, S. Povey and R. Ber (1980) J.N.C.I. *64*, 725−735
4. C.B. Lozzio and B.B. Lozzio (1975) Blood *45*, 321−334
5. P.A. Marks and R.A. Rifkind (1978) Ann.Rev.Biochem. *47*, 419−426
6. G. Stamatoyannopoulos and A.W. Nienhuis (1989) Hemoglobin Switching, Part A. publ. A.R. Liss, New York
7. F. Grosveld, G.B. van Assendelft, D.R. Greaves and G. Kollias (1987) Cell *51*, 975−985
8. R. Treisman, M. Green and T. Maniatis (1980) Proc.Natl.Acad.Sci.USA *80*, 7428−7432
9. E. Whitelaw, P. Hogben, O. Hanscombe and N.J. Proudfoot (1989) Mol.Cell.Biol. *9*, 241−251
10. C. Norman, V. Raymond and N.J. Proudfoot (1987) In: Developmental control of globin gene expression (eds. G. Stamatoyannopoulos and A.W. Nienhuis) pp. 235−252, publ. A.R. Liss, New York
11. N.J. Proudfoot, T.R. Rutherford and G.A. Partington (1984) EMBO J. *3*, 1533−1540

12. M. Plumb, J. Frampton, H. Wainwright, M. Walker, K. Macleod, G. Goodwin and P. Harrison (1988) Nucl.Acids Res. *17*, 73–92
13. T. Evans, M. Reitman and G. Felsenfeld (1988) Proc.Natl.Acad.Sci.USA *85*, 5976–5980
14. D.I.K. Martin, S-F. Tsai and S.H. Orkin (1989) Nature *338*, 435–438
15. G. Superti-Furga, A. Barberis, G. Schaffner and M. Busslinger (1988) EMBO J. *7*, 3099–3107
16. E. deBoer, M. Antoniou, V. Mignotte, L. Wall and F. Grosveld (1988) EMBO J. *7*, 4203–4212
17. T. Maniatis, E.F. Fritsch and J. Sambrook (1982) Molecular Cloning: a Laboratory Manual (Cold Spring Harbor Laboratory, New York)
18. C.M. Gorman, L.F. Moffat and B.H. Howard (1982) Mol.Cell.Biol. *2*, 1044–1051
19. P. Lamb, P. Watt and N.J. Proudfoot (1990) Nucl.Acids Res. in press
20. F.L. Graham and A.J. Van der Eb (1973) Virology *52*, 456–467
21. P. Herbomel, B. Bourachot and M. Yanic (1984) Cell *39*: 653–662
22. D.A. Melton, P.A. Krieg, M.R. Rebagliati and T. Maniatis (1984) Nucl.Acids Res. *12*: 7035–7056
23. J.D. Dignam, R.M. Lebowitz and R.G. Roeder (1983) Nucl.Acids Res. *11*, 1475–1489
24. A. Maxam and W. Gilbert (1980) Meth.Enzymol. *65*, 499–560
25. M.M. Garner and A. Revzin (1981) Nucl.Acids Res. *9*, 3047–3059
26. V. Mignotte, L. Wall, E. deBoer, F. Grosveld and P-H. Romeo (1989) Nucl.Acids Res. *17*, 37–55
27. L.A. Chodosh, J. Olesen, S. Hahn, A.S. Baldwin, L. Guarente and P.A. Sharp (1988) Cell *53*, 25–35
28. R. Mantovani, N. Malgaretti, S. Nicolis, A. Ronchi, B. Giglioni and S. Ottolenghi (1988) *16*, 7783–7797
29. L. Wall, E. deBoer and F. Grosveld (1988) Genes & Dev. *2*, 1089–1100
30. M. Antoniou, E. de Boer, G. Habets and F. Grosveld (1988) EMBO J. *7*, 377–384
31. M. Reitman and G. Felsenfeld (1988) Proc.Natl.Acad.Sci.USA *85*, 6267–6271
32. K.A. Jones, J.T. Kadonaga, P.J. Rosenfeld, T.J. Kelly and R. Tjian (1987) Cell *48*, 79–89
33. B.J. Graves, P.F. Johnson and S.L. McKnight (1986) Cell *44*, 565–576
34. N.J. Proudfoot, A. Gil and T. Maniatis (1982) Cell *31*, 553–563